INTEGRATION OF AUTHENTICATION AUTHORIZATION AND ACCOUNTING SERVICE AND PROXY SERVICE

BACKGROUND OF THE INVENTION

1. Field of the Invention

The present invention relates to the field of data communications networks. More particularly, this invention relates to a method and apparatus for unifying the operation of authentication, authorization and accounting services and proxy services in a data communications network.

2. The Background

ISPs (Internet Service Providers) and Telcos (telephone companies) typically offer wholesale internet access and retail internet access to their subscribers. Wholesale access is typically offered to subsidiary and specialized service providers, CLECs (Competitive Local Exchange Carriers), corporations, and Community of Interest (COI) providers. Naturally, the processing afforded customers of the wholesale variety differs from the processing afforded customers of the retail variety. Subscriber information for individual wholesale users is usually stored by those who lease data communications network access from the ISP or Telco. Hence, corporations, CLECs and COI providers do not normally share their user information with the wholesale providers. The ISP or Telco, however, typically also has its own retail subscribers whose user information is stored in its databases. Hence, the ISP or Telco must identify an incoming user as a wholesale user or a retail user and initiate different actions for an incoming user based upon this status.

See, for example, FIG. 1 where a pure retail environment has a number of network access servers (NAS_1, NAS_2 and NAS_3) which provide data communications portals to the ISP's point of presence (PoP) on the data communications network. Each NAS is in communication with a conventional AAA (authentication, authorization and accounting) service maintained by the ISP. Incoming users connect to the NASes by dialing in over the telephone network or in another conventional manner.

Traditional wholesale ISPs and Roaming Service Providers offer network access through a technique called "Authentication proxying." Proxying involves the transfer of the Authentication responsibility to the "owner" of the subscriber. Thus, if a corporation was to outsource its corporate intranet to an ISP, what it gives up is the maintenance of its dial-up servers (i.e., the NASes). It does not, however, normally want to give up the control or information of its employees. Hence, when a corporate user dials in to such an ISP's network access servers, the user essentially perceives that the user is dialing into a corporate facility when the user is actually dialing into the ISP's domain and then somehow gaining admittance to the corporation's intranet.

What really happens in that scenario is that the ISP determines that the user belongs to Corporation A($Corp_A$) by parsing either the fully qualified domain name (FQDN) supplied by the user, a DNIS ID, or some other mechanism. Having determined that the user trying to gain access belongs to $Corp_A$, the ISP cannot really authenticate the user. As noted earlier, the user's record is still with the corporation. Hence, the ISP will "proxy" out the authentication transaction to the corporation. An AAA service within the corporation then identifies the user, verifies the password, and provisions the user. Then the AAA service notifies the ISP's proxy server that the user is acceptable and passes along provisioning details associated with the user (such as an IP address to use or a pool identification of an IP address pool from which an IP address needs to be allocated). The ISP then grants the user access to the network based upon the reply it gets back from the corporation. This technique is called "proxying." This is shown in FIG. 2.

To be able to do this, the ISP maintains minimal information on its proxy server **14** at its PoP. Information such as supported domain names, the IP address to which the transaction is to be sent, the port number to which the transaction is to be addressed, etc. are stored (see FIG. 3).

For example, turning now to FIG. 2, user Joe@corpa.com dials in 40 to NAS_1. A PPP (point to point protocol) session is raised between Joe and NAS_1. An IPCP (internet protocol control protocol) session **42** is raised between NAS_1, and proxy service **44**. In response NAS_1 sends a RADIUS (Remote Authentication Dial-In User Service protocol) access-request to proxy service **44**. Proxy service **44** then consults its local configuration database **16**. Proxy service **44** then makes a determination about where to send the access-request packet. Here it decides to send it to the AAA service **48** maintained in the $Corp_A$ domain **50**. The $Corp_A$ AAA **48** then consults its local database **52** and authenticates joe@corpa.com. $Corp_A$ AAA **48** then returns an access-accept packet to proxy service **44** which, in turn, sends an access-accept packet to NAS_1 completing the log-in of joe@corpa.com.

When the subscriber is granted access, or leaves the network, the accounting transactions will now have to be shared with the wholesale customers of the ISP/Telco. That is, the ISP/Telco will keep a record with which to bill or otherwise account to $CorP_A$ for services rendered and the record will also need to be sent to $Corp_A$'s AAA. Typically, the wholesale provider (e.g., the ISP) will use a roaming service product such as the Global Roaming Server™ (GRS), a product of Cisco Systems, Inc. of San Jose, Calif., to achieve this objective. In the retail case, the ISP/Telco will use a product like Cisco Secure™, a product of Cisco Systems, Inc., to act as an authentication, authorization and accounting (AAA) service to authenticate and authorize the user. This approach, however, poses some problems for the ISP/Telco.

The ISP/Telco needs to maintain two different sets of NASes as diagrammed in FIG. 4 or it has to pipe all transactions through a GRS (proxy service) as diagrammed in FIG. 5 which then has to make a decision as to whether the access-request transaction will be locally processed by the ISP/Telco (retail user) or remotely processed by the wholesale customer (wholesale user). The two products are independent products which maintain their own databases. They do not at present support a distributed architecture and hence will not scale by the number of PoPs users, etc. This poses the problem that multiple instantiations of the GRS will need to be configured and will not be able to properly load balance among the various NASes available at the PoP. Furthermore, should a GRS go down, the PoP may lose the services of the NASes in communication with the GRS that failed.

Accordingly, it would be desirable to provide a capability for allowing ISPs and Telcos to seamlessly offer wholesale and retail data communications network access, unify the disparate systems that specialize in these access control segments and scale both systems to simultaneously reside on a plurality of PoPs while behaving in a distributed manner within the data communications network.

SUMMARY OF THE INVENTION

A single database maintained centrally hosts both proxy service data and authentication, authorization and account-

ing (AAA) data. Data is then copied to storage used locally by each system when both systems are instantiated. Therefore the ISP/Telco need not maintain two different data bases. A protocol gateway (PGW) is used to determine if the incoming user is a wholesale or retail user. The PGW filters the domain portion of the access request to locate a remote AAA service. If one such service is found, the PGW routes the communication via the GRS to proxy it to the remote AAA service. The returned packet from the remote AAA service is then searched for an IP address to be assigned to the incoming user. If one is not found the PGW obtains a dynamically allocated IP address from a DHCP server (using an IP-Pool-ID if supplied in the returned packet from the remote AAA service). The same mechanism is used to forward accounting event packets from the NAS to the remote AAA service. The PGW may monitor more than one proxy service and/or AAA service and load balance among them.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 is a system block diagram of a simple ISP PoP using a conventional retail-only paradigm.

FIG. 2 is a system block diagram of wholesale ISP PoP using a conventional wholesale-only paradigm.

FIG. 3 is a diagram illustrating the information maintained by a conventional proxy server.

FIG. 4 is a system block diagram of an ISP PoP having non-integrated retail and wholesale components.

FIG. 5 is a system block diagram of an ISP PoP using a Global Roaming Server (GRS) proxy service to integrate wholesale and retail functions.

FIG. 6 is a system block diagram of an ISP PoP using a protocol gateway (PGW) in accordance with a presently preferred embodiment of the present invention to integrate wholesale and retail functions and perform load balancing.

FIG. 7 is a system block diagram of an ISP NOC, broker publisher system and PoP in accordance with another preferred embodiment of the present invention.

FIG. 8 is a system block diagram of a broker publisher system used in accordance with a preferred embodiment of the present invention.

FIG. 9 is a flow diagram detailing a process by which the AAA service and its associated database are instantiated in accordance with a presently preferred embodiment of the present invention.

FIG. 10 is a flow diagram detailing a process by which a proxy service and its associated database are instantiated in accordance with a presently preferred embodiment of the present invention.

FIG. 11 is a flow diagram detailing a user authentication and authorization process in accordance with a presently preferred embodiment of the present invention.

FIG. 12 is a flow diagram detailing a load balancing process in accordance with a presently preferred embodiment of the present invention.

FIG. 13 is a flow diagram detailing an accounting process in accordance with a presently preferred embodiment of the present invention.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

Those of ordinary skill in the art will realize that the following description of the present invention is illustrative only and not in any way limiting. Other embodiments of the invention will readily suggest themselves to such skilled persons after a perusal of the within disclosure.

In accordance with a presently preferred embodiment of the present invention, the components, processes and/or data structures are implemented using a gateway device and other services implemented using C++ programs running on an Enterprise 2000™ server running Sun Solaris™ as its operating system. The Enterprise 2000™ server and Sun Solaris™ operating system are products available from Sun Microsystems, Inc. of Mountain View, Calif. Different implementations may be used and may include other types of operating systems, computing platforms, computer programs, firmware and/or general purpose machines. In addition, those of ordinary skill in the art will readily recognize that devices of a less general purpose nature, such as hardwired devices, devices relying on FPGA (field programmable gate array) or ASIC (Application Specific Integrated Circuit) technology, or the like, may also be used without departing from the scope and spirit of the inventive concepts disclosed herein.

The protocol gateway (PGW or gateway) is a device which couples the user via a network access server (NAS) to the data communications network. The term gateway is not meant to be limited to a single type of device, as any device, hardware or software, that may act as a bridge between the user and the network may be considered a gateway for the purposes of this application. In accordance with a presently preferred embodiment of the present invention, the PGW is a software service operating on a general purpose computer running the User Control Point (UCP) software package available from Cisco Systems, Inc. of San Jose, Calif.

The authentication, authorization and accounting (AAA) service performs user authentication, user authorization and user accounting functions. It may be a Cisco ACSTM produce such as Cisco Secure™, available from Cisco Systems, Inc. of San Jose, Calif., or an equivalent product. In accordance with a presently preferred embodiment of the present invention, the Remote Authentication Dial-In User Service (RADIUS) protocol is used as the communication protocol between the gateway and the AAA and GRS proxy services. RADIUS is an Internet standard track protocol for carrying authentication, authorization, accounting and configuration information between devices that desire to authenticate their links and a shared AAA or GRS service. Those of ordinary skill in the art will realize that other Internet protocols such as TACACS+ can be used as acceptable authentication communications links between the various communications devices that encompass the data communications network and still be within the inventive concepts disclosed herein. The global roaming service (GRS) is also a AAA service which is capable of proxying transactions to remote AAA service. It also preferably uses the RADIUS protocol or an equivalent.

One way in which the present invention may come into use involves the concept of roaming users. A roaming user is, for example, a traveling person with a lap top. If the person wants to reach a corporate intranet or local ISP, he or she can (1) dial the number of the home PoP (point of presence) and incur potentially large telephone bills; (2) dial a "toll free" number such as an 800 number which can also be expensive—to the provider; or (3) use a global roaming server model. In the global roaming server model, ISPs with PoPs in different locations make cross-agreements with one another so as to provide local telephone access numbers to ISPs without any other (or a sufficient) presence in a location. To the user, it appears that his ISP has PoPs everywhere that there is a roaming agreement in place with a cooperating ISP.

A global roaming service ("GRS") at a PoP can parse the fully qualified domain name ("FQDN") of the user (e.g., joe@ISPA.NET) and determine that Joe belongs to ISPA-.NET. The GRS can then send an authentication request to ISPA.NET's AAA server to authenticate and authorize Joe in a conventional manner. Accounting event information, e.g., accounting start packets associated with log-in and accounting stop packets associated with log-out, are sent both to the GRS at the local PoP and to ISPA.NET's AAA server to enable the local PoP to account for use by Joe at the local PoP and so bill ISPA.NET, if desired, and to allow ISPA-.NET to bill Joe, if desired. It also provides a mechanism for tracking this type of usage which can serve a number of purposes.

GRSes have their own associated databases which keep lists of remote AAAs, their IP addresses, their port numbers and their associated domain names.

To render the roaming model more tenable to the myriad IPSs and Telcos which might see fit to enter into these cross-agreements and thus make roaming easier for the end users, the process must be simplified and made scaleable. Under the prior model, as shown in FIG. **6**, each GRS and AAA had its own associated stand-alone database which required maintenance from time to time. Multiple instances of such databases required individual maintenance. In many situations NAS resources were committed to a particular AAA or GRS at a PoP and not capable of load balancing.

FIG. **7** is a system block diagram of an improved system in accordance with a presently preferred embodiment of the present invention. A data communications network **10** such as the internet, or an ISPs presence on the internet, or a corporate intranet, or the like, includes a network control console (NCC) **12** which is physically located on a host **14** within a Network Operations Center (NOC) **16**. The NCC **12** is an application running on the host **14**. The NCC **12** monitors and manages the data communications system. The NCC **12** is in communication with a database **18** and an access database adapter **20**.

The database **18** and access database adapter **20** can run on the same host **14** as the NCC **12**, as depicted in FIG. **7**, or the database **18** and the access database adapter **20** can be located on more than one device. The database **18** stores information related to the various components and services comprising the data communications network **10** being managed. The system administrator accesses the information in the database **18**, as needed, in conjunction with the NCC **12**, to perform the overall network management task. The access database adapter **20** is in communication with both the database **18** and the NCC **12**. This adapter, and other adapters in the invention, provide bi-directional mapping of information between the NCC **12** and other services comprising the data communications network **10**. Adapters, such as the access database adapter **20** subscribe to and publish events. An event is an independent entity which contains an unspecified amount of non-time critical information. For example, the access database adapter **20** receives commands from the NCC **12** to publish an event. The information contained in the event may be found in the NCC's request or the access database adapter **20** may communicate with the database **18** to find the required information. A detailed discussion of some of the specific events pertinent to this invention and the information found therein is provided later in this disclosure. The event is then published to other services and components within the data network management system across an information bus **22** which may be the data communications network itself.

The information bus **22** that serves as the transportation medium for the presently preferred embodiment of the present invention can be Common Object Request Broker Architecture (CORBA)-based. The CORBA-based information bus is capable of handling the communication of events to and from objects in a distributed, multi-platform environment. The concept of a CORBA-based information bus is well known by those of ordinary skill in the art. Other acceptable communication languages can be used as are also known by those of ordinary skill in the art.

CORBA provides a standard way of executing program modules in a distributed environment. A broker **24**, therefore, may be incorporated into an Object Request Broker (ORB) within a CORBA compliant network. To make a request of an ORB, a client may use a dynamic invocation interface (which is a standard interface which is independent of the target object's interface) or an Object Management Group Interface Definition Language (OMG IDL) stub (the specific stub depending on the interface of the target object). For some functions, the client may also directly interact with the ORB. The object is then invoked. When an invocation occurs, the ORB core arranges so a call is made to the appropriate method of the implementation. A parameter to that method specifies the object being invoked, which the method can use to locate the data for the object. When the method is complete, it returns, causing output parameters or exception results to be transmitted back to the client.

In accordance with a presently preferred embodiment of the present invention an Enterprise Application Integration (EAI) system is used to broker the flow of information between the various services and adapters comprising the data network management system of the present invention. An example of an EAI system that can be incorporated in the presently preferred invention is the ActiveWorks Integration System, available from Active Software of Santa Clara, Calif. As shown in FIG. **8**, such an EAI system **26** uses an information broker **24** as the hub of the system. The information broker **24** acts as the central control and storage point for the system. The information broker **24** can reside on a server and serves to mediate requests to and from networked clients; automatically queuing, filtering and routing events while guaranteeing delivery. The information broker **24** is capable of storing subscription information and using such subscription information to determine where published information is to be sent. Referring back to FIG. **7**, the information broker **24** is shown as being located at a point along the information bus **22**. In most instances the, broker will be located within the same NOC **16** as the host **14** that runs the NCC **12** application. Another key feature to the EAI system **26** of FIG. **8** is the use of adapters **28***a*, **28***b*, and **28***c* that allow users of the EAI system **26** to integrate diverse applications and other information when using the integration system. Adapters **28***a*, **28***b*, and **28***c* provide bi-directional mapping of information between an application's native format and integration system events, enabling all custom and packaged applications, databases, and Internet and other network applications to exchange information. As shown in FIG. **8** the adapters **28***a*, **28***b*, and **28***c* run in association with the various services **30***a*, **30***b*, and **30***c* from which information is published and subscribed on to an information bus **22** that has its hub at the broker **24**.

Referring back to FIG. **7** the information bus **22** is in communication with a Point of Presence (POP) **32** within the data communications network **10**. The PoP **32** is one of many PoPs that the information bus **22** is in communication with. Located within PoP **32** is a host or node **34** which may comprise one or more computing devices on which some or all of the services shown in FIG. **7** may be running. The node

34 is in communication with the information bus 22 through a control adapter 29 which provides control communications with the various services 30*a*, 30*b*, 30*c*, 30*d*, 30*e* through their respective service adapters 28*a*, 28*b*, 28*c*, 28*d*, 28*e* via service adapter 31 of control adapter 29.

By way of example, the node 34 of FIG. 7 is configured with a PGW 30*a*, an authentication, authorization and accounting (AAA) service 30*c*, a domain name system (DNS) service 30*e*, a dynamic host configuration protocol (DHCP) service 30*d* and a pair of GRS services 30*b*. Those of ordinary skill in the art will appreciate that the services shown are not intended to be limiting and that other services and other service configurations can be used without departing from the inventive concepts herein disclosed. The system services may also be distributed over two or more servers to provide improved performance and redundancy.

The protocol gateway service 30*a* is used to couple the network user to the data communication network. The protocol gateway service 30*a* functions as an interface to the NASes that allows access requests received from a user to be serviced using components that may communicate using different protocols. A typical protocol gateway service 30*a* may be able to support different user access methodologies, such as dial-up, frame relay, leased lines, ATM (Asynchronous Transfer Mode), ADSL (Asymmetric Digital Subscriber Line) and the like. Used in conjunction with the protocol gateway service 30*a*, the AAA service 30*c* performs user authentication, authorization and accounting functions. The AAA service 30*c* stores user profile information and tracks user usage. The profile information stored in the AAA service 30*c* is proxied to the protocol gateway service 30*a* when a network user desires network access.

The DNS service 30*e* is used to return Internet protocol (IP) addresses in response to domain names received, for example, from a protocol gateway service 30*a*. For example, if the DNS service 30*e* receives a domain name query from the protocol gateway service 30*a*, it has the capability to locate the associated numerical IP address from within the memory of the DNS service (or another DNS service) and return this numerical IP address to the protocol gateway service 30*a*.

The DHCP service 30*d* is used as a dynamic way of assigning IP addresses to the network users as well known to those of ordinary skill in the art.

Each of these services 30*a*, 30*b*, 30*c*, 30*d*, 30*e* is in communication with a corresponding service adapter 28*a*, 28*b*, 28*c*, 28*d*, 28*e*. The service adapter subscribes to and publishes various events on the information bus 22. The service adapter is configured so that it subscribes to events published by the access database adapter 20 of the NCC 12. The service adapter also publishes events to the access database adapter 20 of the NCC 12.

The following is an exemplary listing and definition of some of the events published by and subscribed to by the access database adapter and the service adapters which are pertinent to this invention. This listing is by way of example and is not intended to be exhaustive or limiting in any way. Other events are possible and can be used in this invention without departing from the inventive concepts herein disclosed.

The NCC 12 publishes "configure" events to the service adapters 28*a*, 28*b*, 28*c*, 28*d*, 28*e*. Configure events are published to configure the service adapters upon initial start up of the service adapters or to modify a preexisting configuration. A configure event can be delivered to a service adapter directly from the access database adapter 20 at the NCC 12. The service adapters update their corresponding configuration files upon receiving a configure event. An example of the information contained within a configure event includes the GUID (global unique identifier) of the publisher, the GUID of the subscriber, listening port configuration, sink port configuration, protocol handler information, engine data and facility data.

The NCC 12 publishes "start" events that are subscribed to by a control adapter such as control adapter 29 associated with a host computer at a node to cause the control adapter to start up one or more specific services. Since the control adapter is always responsible for starting a service, the start events are always subscribed to by the control adapters as opposed to the service adapters. An example of the information contained within a start event includes the GUID of the publisher, the GUID of the subscribing control adapter, the GUID of the service to be started, the service name and the absolute path where the service binary resides. The access database adapter 20 of the NCC 12 also publishes "stop" events that are subscribed to by the control adapter to cause the control adapter to shut down a specific service or multiple services. Since the control adapter is always responsible for stopping a service, the stop events are always subscribed to by the control adapter as opposed to the service adapters. Once the control adapter receives the stop event, it publishes a stop event to the service adapter of the corresponding service. The control adapter allows the service sufficient time to shut down. If the service does not respond to the stop event and continues running the control adapter can explicitly kill the service based on the process ID found in the configuration file. An example of information contained within a start event includes the GUID of the publisher, the GUID of the subscribing control adapter, the GUID of the service to be stopped and the name of the service to be stopped.

Other events may be published and subscribed to.

The configure event is used to publish the current contents of a master database relevant to GRS and AAA services at the various nodes of the data communications network. Thus the master database may be maintained and serviced at the NOC or some other convenient facility and the AAA services and GRS services updated with information automatically without the need to manually update their separate databases.

The PGW is used as a protocol gateway between the NASes and the AAA and GRS services. The PGW parses the FQDN of incoming users and sends access requests from local users to the local AAA and access requests for roaming users to the GRS. The GRS, in turn, forwards the access requests to the remote AAA belonging to the user's provider in accordance with the conventional proxy model.

The PGW has the ability to load balance by monitoring the condition and response times of its respective GRS services and AAA services. Thus, if one such services is particularly loaded, incoming calls may be directed to other services. If one such server has crashed or becomes non responsive, it may be bypassed. In the present configurations where NASes are directly connected to a GRS and or an AAA service, a dead service can result in the NASes connected to the dead service becoming non-responsive. This condition is avoided by using the PGW as a front end to the GRS and AAA service.

In accordance with the present invention IP addresses may be assigned to incoming users in a number of ways. For users having permanently or otherwise allocated IP addresses reflected in their user service profiles in their